

张心亮

意向岗位 (北京)：多模态大模型 / 算法实习生

☎ 17860713361 · ✉ zhangxinliang@stu.pku.edu.cn · 🌐 zx119990529

🎓 教育背景

北京大学 / 人工智能与计算机视觉方向 / 博士 2025.09 – 至今
天津大学 / 计算机科学与技术 / 硕士 (保研) 2021.09 – 2024.05
中国海洋大学 / 电子信息工程 / 本科 2017.09 – 2021.5

👨‍🔬 研究经历

多模态大模型视觉 token 压缩 2024 – 2025

项目背景：多模态大模型在视觉理解任务中表现出色，但图像平方级别的 token 数量导致计算资源消耗巨大。通过压缩视觉 token 降低复杂度，提高 LLM 的推理速度，从而为模型整体节约成本。

工作内容：提出了一种 object-level 的视觉 token 压缩的方法，设计并实现了一个高效的视觉 token 压缩模块，能够根据输入图像内容自适应地选择 token 进行保留，从而大幅减少视觉 token 数量。

工作成果：1) 基于 LLaVA, InternVL 架构在仅用 10% 的视觉 token 的情况下，能保留 96% 的性能，同时显著降低模型的幻觉。在单张 4090 上，全流程解码 128tokens 的平均时间比原始模型减少 25%。2) 视觉压缩模块与原始 LLM 解耦，可以解耦式部署，降低对数据上传的带宽需求，提升用户隐私保护。

基于稀疏标注学习的语义分割方法 2023 – 2024

项目背景：高精度标注需要大量人力成本，稀疏标注廉价易得，但带来的监督信号薄弱。如果能在稀疏标注的数据下提升语义分割模型的性能，可以大大降低对高精度数据的依赖，从而降低标注成本。

工作内容：1) 提出了一种基于稀疏标注的语义分割方法，通过引入类别监督生成伪标签和局部特征矫正模块，有效提升了模型在有限标注数据下的性能，大大降低了标注数据的成本。2) 提出了一个模拟人工偏好标注数据集构建策略，制作了一个大规模稀疏标注语义分割数据集。

工作成果：1) 在多个公开数据集上验证了所提方法的有效性，在 Pascal VOC、COCO 等数据集上，使用 10% 的涂鸦标注数据，达到了与全监督方法相当的性能。2) 在大尺度、复杂场景下的医学图像、自动驾驶等分割任务中，达到了 SOTA 性能。3) 揭示了标注规模可以与标注精度互补的关系。

🏠 学术成果

[google scholar](#)

AdaTok: Adaptive Token Compression with Object-Aware Representations for Efficient Multimodal LLMs. *CVPR'26* (一作, 在投) 提出了自适应视觉 token 压缩方法，保持高性能的同时显著提升推理效率。

Promoting Scribble-Based Weakly-Supervised Semantic Segmentation with Its Class Label. *AAAI'24* (一作, *CCF-A*) 提出了融合类别标签的涂鸦监督语义分割，利用稀疏标签实现高精度分割，降低标注成本。

Generative learning-based lightweight MRI brain tumor segmentation with missing modalities. *ESWA* (一作, 一区 *TOP*) 提出了基于生成学习的轻量级 MRI 脑肿瘤分割方法，能够有效处理缺失模态问题。

Scribble Supervised Multimodal Medical Image Segmentation. *IJCNN'24* (导师一作, *oral*)

学术兼职：担任 *AAAI*、*CVPR* 等会议审稿人。担任 *IEEE TPAMI*、*TIP*、*TMI* 等期刊审稿人。

🔧 技术栈 & 专业技能

[MILab](#)

- (存储) 基于 BeeGFS 和 IB 拓扑搭建高性能分布式文件系统，解决高并发情况下的 IO 瓶颈问题。
- (网络) 基于 NAS 和 NFS 文件系统，端配置共享工作目录，配置 pip/anaconda3、huggingface 公共缓存，通过指定缓存路径实现资源统一复用，大幅降低成员依赖包下载耗时，提升研发效率。
- (Web) 基于 Confluence 搭建组内知识库，实现技术文档的统一管理与分享，提升团队协作效率。
- (管理) 基于 Python+Flask 搭建 GPU 监控网站，实现对集群用户使用情况的实时监控与统计。